Interactive system of terminological dictionaries as one of the elements in the ontology of scientific knowledge

N.E. Kalenov [1]*, Dr. Sc. (Engineering), Professor, Chief Researcher, nekalenov@yandex.ru*
A.M. Senko [2]*, Researcher System Research & Automation Technology Dpt.,*
alexander.senko@gmail.com

[1] *Joint Supercomputer Centre of RAS, Moscow, 119334, Russia*
[2] *Library for Natural Sciences of RAS, Moscow, 119991, Russia*

The authors consider the interactive informational system. It is in the Internet. It has over 12, 000 terms belonging to 69 thematic sections (dictionaries) of science and technics. They fit to the level of the State Rubricator of Scientific and Technical Information. Each term has its definitions (probably several) and links to the definition sources. The system includes rubricators of universal decimal classification and library-bibliographical classification (partially). Also the authors defined the correspondence of indexes of these classification systems. They are in the dictionaries. Besides that, authors present the connections between the terms both within one dictionary and with the terms of other dictionaries.

The system orients to the decentralized support and distributed data editing and refilling. In the system, there are three foreseen levels of access rights: 1) user access with rights for data search, data view and developed navigation; 2) editor access, where there are also the rights for entering and editing data in the user rights; 3) administrator access, where the possibility to add new users and set their rights, and modification of the database structure are added to the editor rights. The system allows viewing the information update history (what, when and by who changed something inside the content) as well as ensures a developed navigation among its own elements.

The further development of the system is its integration into the model of the common digital space of scientific knowledge developed with the support of RFBR as a basis for forming the ontology of scientific directions that belongs to the core of this space.

Keywords: *terminological dictionaries, classification systems, ontology of scientific knowledge, scientific space, scientific databases, interactive system, Internet.*

The pilot version of the interactive *system of terminological dictionaries* (STD) became a basis for creating the ontologies describing the subject areas of science [1-3]. It was developed in 2018 within the project supported by Russian Humanitarian Science Foundation. The project provided for the download of terminological dictionaries prepared by experts of the All-Russian Institute of Scientific and Technical Information (ARISTI) into the relational *database (DB)* as a set of MS-Word files [3]. Besides that, they planned a development of an interface that would provide the possibility to edit, enter new data and to navigate through the STD resources [4-6].

As a result of the project realization, an interactive STD appeared that contained information from 69 terminological dictionaries; each of them fits to one of the upper-level sections of the SRSTI - *the State Rubricator of Scientific and Technical Information* [6]. The pilot project of STD database contained the terms from a dictionary for a matching science discipline, the definitions of terms and connected sources. The definition are from them.

The authors developed the system on the basis of SciRus software package, which is a basis for the technological block of the electronic library "Scientific Heritage of Russia" and a range of other informational systems [7, 8]. The core of the system is a relational database; MySQL 5.0 is as DBMS. During the system creation, the authors used the frame of Ruby on Rails web apps in version 4.1.

The works on the STD development goes on. They are developed in two directions:

· Expansion of the informational system base by including indexes of other classification system in it – *universal decimal classification* (UDC) and *library-bibliographical classification* (LBC);

· Establishment of thesaurus connections between dictionary terms and various classification indexes.

The authors suppose that in the future, STD complemented with keywords from thematic databases will serve as a basis for creating ontologies of scientific directions and will be built into the common digital space of the scientific knowledge.

There are the characteristic of STD working at the moment and examples of working with the system below.

STD Structure

There is the developed STD on the server of *the Library for Natural Sciences of the Russian Academy of Sciences (RAS LNS)* at *http://class.labs.benran.ru/*. It operates with the following object types: users, dictionaries, terms, definitions of terms, their sources, UDC indexes, LBC indexes, connections between the terms.

The "user" object is described by the fields: username (login); password to enter the system; user category (visitor, editor, administrator). A user with the status "visitor" has the right to search and view the information; the "editor" status provides the right to input and edit data; the "administrator" status gives the right to change user rights, enter new users, and add new fields to the objects' descriptions.

The "dictionary" object has the mandatory fields "name" and "code" (upper-level index of the SRSTI); each dictionary is connected by the ratio 1 : n with the terms that belong to this science section.

The "term" object has a mandatory field "name", an optional field "index of the SRSTI", a mandatory connection of the type (1 : n) with the definitions of terms (each term must have at least one definition, but they may be several of them), optional connection with the UDC and LBC indexes and also optional references to the connections with other terms of the type (n : n).

The "term definition" object has a field that contains a definition text and a connection of the type (n : 1) with the "definition source" object.

The "definition source" object contains text fields "name" and "additional information" and connections of the type (1 : n) with the "term definition" objects.

The "connection" object designates the connections between a pair of terms and contains a link to the first term, a link to the second term and the connection type between the first and the second term, which may accept one of four values: "equal", "belongs to", "contains", and "intersects".

The "UDC" and "LBC" objects contain the fields "section name", "section index" and optional connections of the (n : n) type with the terms.

Work with the system

After the authorization, a user comes to the first system page (fig. 1).

Each dictionary name is an active link; by following it, there is an opening window with a message about the number of terms that belong to this dictionary.
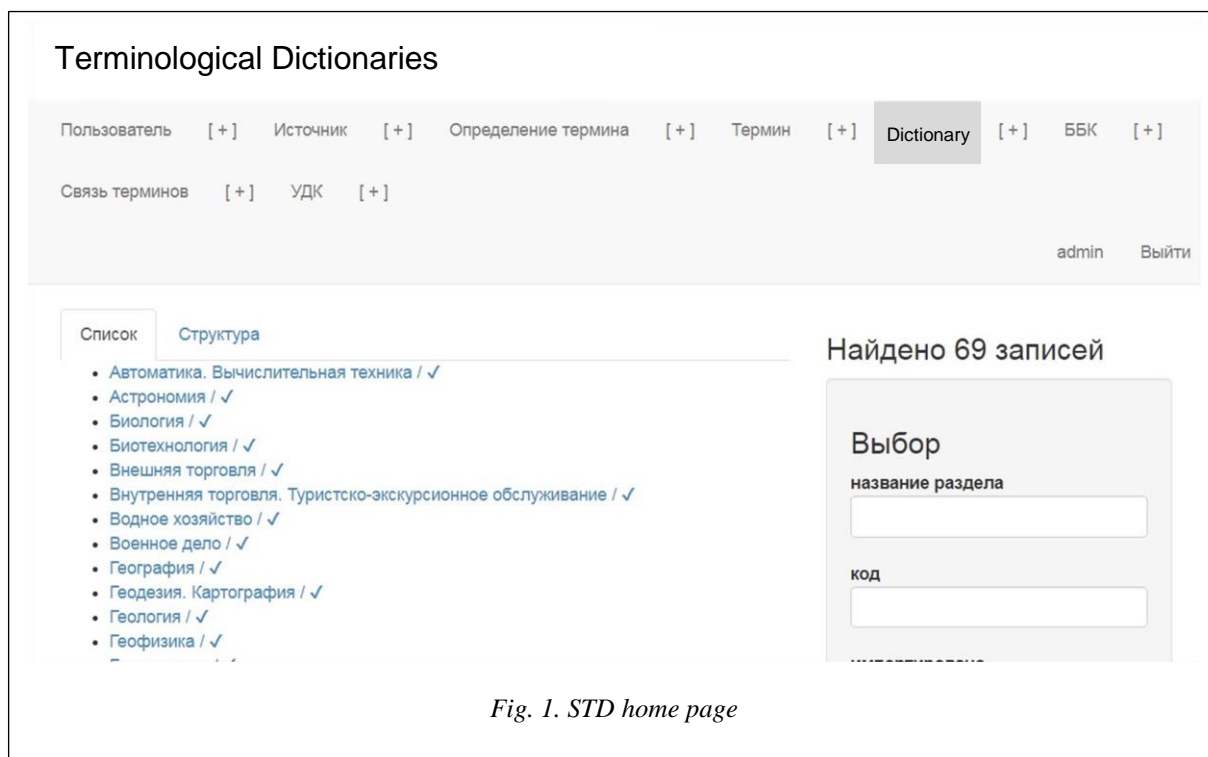

Fig. 1. STD home page

The example on figure 2 shows the number of terms that belong to the dictionary "Automation and computer engineering" – there are 170 of them in total.

By following the "View" link, a user can see an alphabetical list of corresponding terms (fig. 3). Each term from the list is an active link that opens a window with detailed information about this term.

Figure 4 shows the page that belongs to the term "algorithms" of the dictionary "Automation. Computer engineering." The page contains the term name with the note, to which dictionary it belongs (with the possibility to go to the dictionary), the indexes of the SRSTI, UDC and LBC that correspond to this term; a link to the term definition (and to the definition source), and a reference to the connection of this term with others.

In the considered example, following the link of the name of the UDC section ("modelling…") opens the page shown on figure 5. It reflects the name and code of the UDC section, and also the link of the list of the STD

terms those belong to this section. In this case the list shown on figure 6 includes 4 terms. Each element of the list is an active link directing to this term's page.
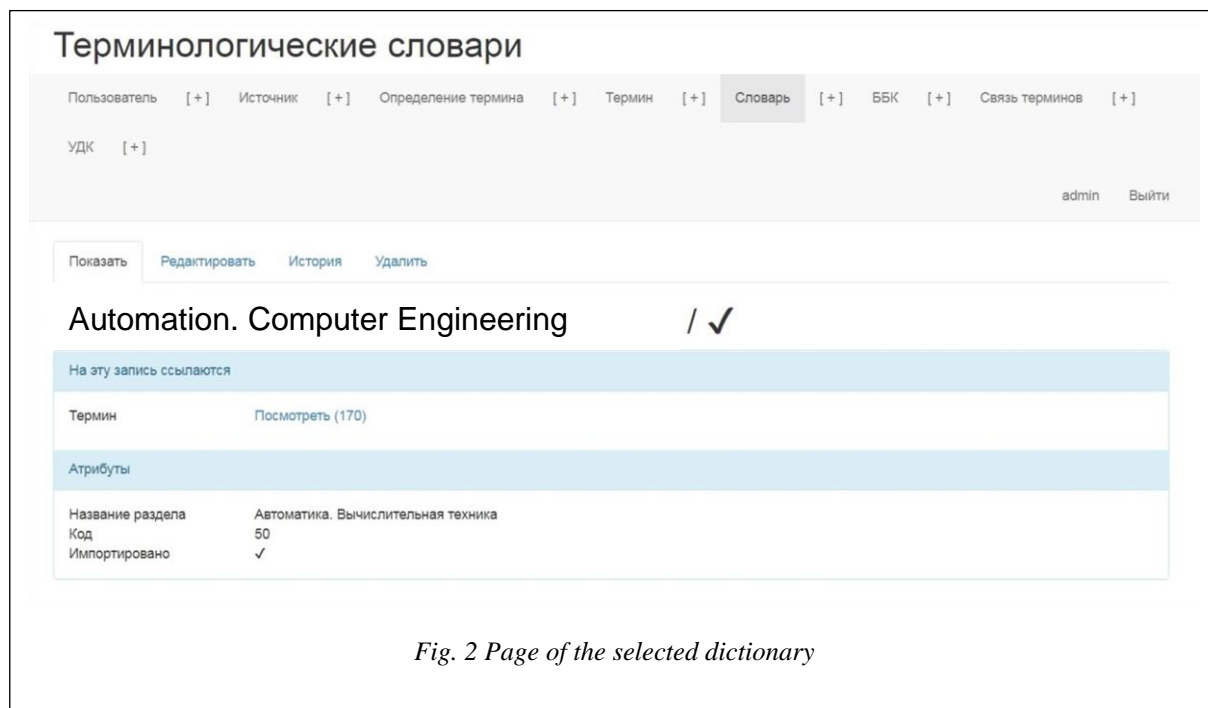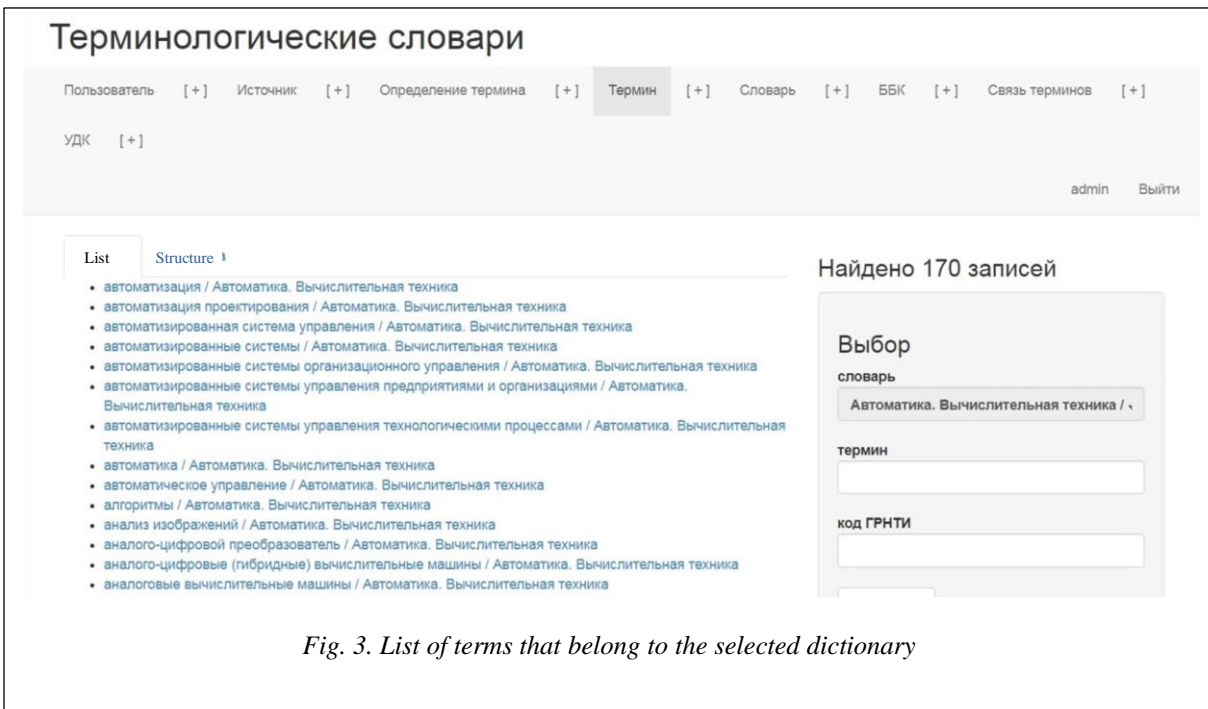

Fig. 2 Page of the selected dictionary

In the considered example, following the link of the name of the UDC section ("modelling…") opens the page shown on figure 5. It reflects the name and code of the UDC section, and also the link of the list of the STD terms those belong to this section. In this case the list shown on figure 6 includes 4 terms. Each element of the list is an active link directing to this term's page.

Fig. 3. List of terms that belong to the selected dictionary

Enlarge upon the page of the term "algorithms" (fig. 4). Besides the "Show" tab, it has the tabs "History", "Edit", "Delete". A click on the "History" tab opens a page showing all operations connected to the changes of this term (what and when changed, names of operator who changed the entry). The operators with the "editor" status got the greenlight to a click on the tabs "Edit" and "Delete". A term deletion is possible only if there are no connections between this term and others.

Clicking the "Show" link in the "Term connection" line opens a page with the list that includes 54 terms connected with the term "algorithms" (fig. 7). In setting of preliminary connections between the STD terms, we used an automatic algorithm that revealed the entry of this term to the definitions of other terms. The operators corrected the connections revealed on this stage – they either specified them by providing one of four values mentioned above or removed. As the figure shows, the term "informational requests" is connected with the terms that belong to various dictionaries.

Fig. 4. Page of the term "algorithms" of the dictionary "Automation. Computer Engineering"

Fig. 5. Page of the UDC section to which the term "algorithm" belongs

Each connection presented in the list is an active link by following which we can see, what connection type exactly ("contains", "belongs to…", "intersects", "equal") is established between these terms. So, by clicking the connection "algorithms - automatic translation" we can come to the page containing a definition of the term "automatic translation" with the statement of "intersects" connection type (fig. 8).

The STD interface allows searching for a necessary object of any type by a fragment of its name. In this regard it's necessary to select a corresponding tab (object type) in the upper part of the home page (fig. 1) and input

a needed fragment in the right part of the page. For the classification systems (UDC and LBC), there is a search by category codes. As shown by the example cases, the system provides a developed navigation through its vari-
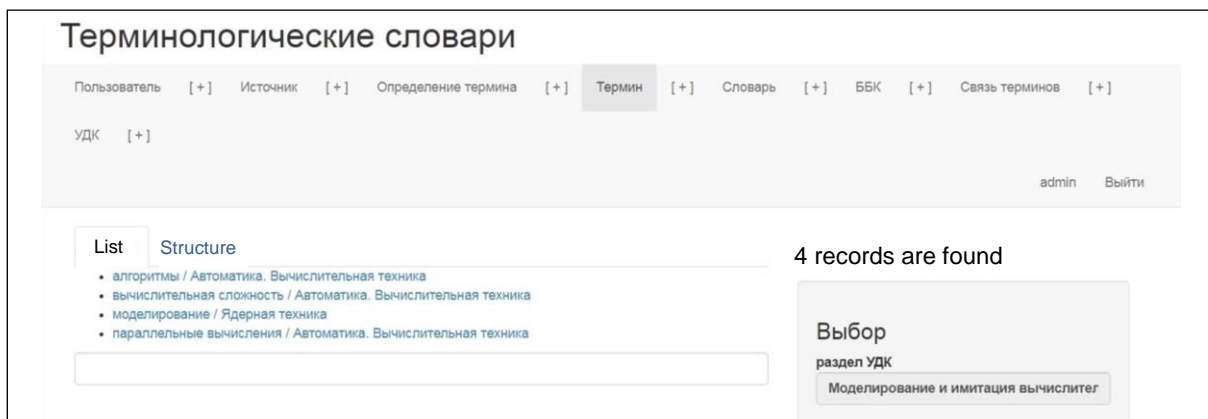
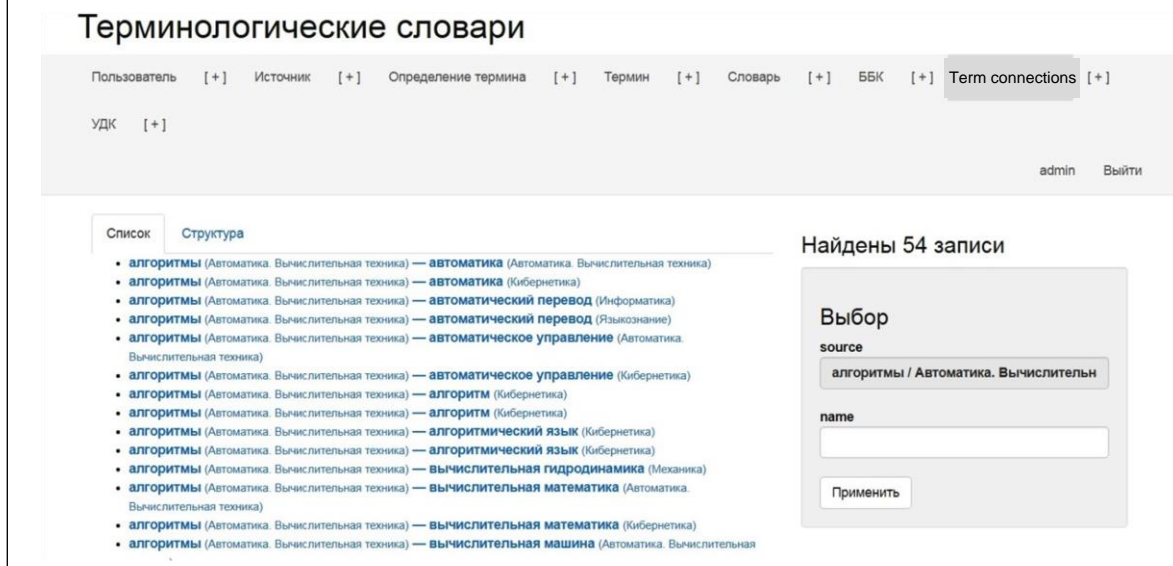*Fig. 6. List of the terms which belong to a specific UDC section*

Fig. 7 Connections of the term "algorithms"

Fig. 8. Connection of terms "algorithms" and "automatic translation"

ous elements; provides full information about the whole change history of its content (which is vitally important for a distributed approach to the STD replenishment).

For the time being, in the STD there are 12, 090 terms that belong to 69 dictionaries; 12, 842 definitions of terms taken from 7, 661 sources; 298, 381 connections among the terms; 3, 025 UDC indexes; 559 LBC indexes.

Conclusion

The developed version of the STD can be used already now to get reference information and formulate detailed queries to the informational search system that include keywords and classification indexes of the SRSTI, UDC and LBC. However, the main purpose of the STD in the long view is to be a basis for building ontologies of separate scientific directions and interdisciplinary subject ontology that are included in the common digital space of the scientific knowledge [9]. A model of such space is currently being developed with the efforts of the experts from the RAS Interdepartmental Center for Supercomputing, the RAS Federal Research Center 'Informatics and Control', the RAS LNS, and the RAS Institute of Scientific Information for Social Sciences.

Refrerenses

1. Beloozerov V.N. The technology for the development of terminological dictionaries with the classification systems vocabulary. *Proc. Informational Support of Science: New Technologies*, 2015, pp. 126-136 (on Russ.).

2. Antoshkova O.A., Beloozerov V.N., Dmitrieva E.Yu., Shapkin A.V. The development of the STI based on bibliographical classifications. *Proc. Informational Support of Science: New Technologies*, 2017, pp. 292-300 (on Russ.).

3. Antopol'skii A.B., Beloozerov V.N., Markarova T.S. About the ontology development based on classifiers of the scientific information and terminological dictionaries. *Informational resources of Russia*. 2017, no. 5, pp. 2-7 (on Russ.).

4. Antopol'skii A.B., Beloozerov V.N., Kalenov N., Shaburova N.N., Yakshin M.M. The development of a semantic network of keywords based on definitive relationships. *Scientific and Technical Information Processing*, 2017, vol. 44, no. 4, pp. 261-265. DOI: 10.3103/S0147688217040062.

5. Antopol'skii A.B., Beloozerov V.N., Kalenov N.E., Markarova T.S. About the development of the terminological database as a complex of industry-specific informational search thesauri. *Informational resources of Russia*, 2018, no. 5, pp. 22-30 (on Russ.).

6. State rubricator of scientific and technical information. Available at: http://www.extech.ru/info/catalogs/grnti/ (accessed June 18, 2019) (on Russ.).

7. Yakshin M.M. The development of the SciRus platform. *Proc. Informational Support of Science: New Technologies*, 2015, pp. 203-207 (on Russ.).

8. SciRus Platform – a basis of the technological complex of the electronic library "Scientific Heritage of Russia". *Proc. RCDL-2014: Electronic Libraries: Promising Methods and Technologies, Electronic Collections*, 2014, pp. 362-368 (on Russ.).

9. *Ruby on Rails 4.1. Release Notes*. Available at: https://edgeguides.rubyonrails.org/4_1_release_notes.html (accessed June 18, 2019).

10. Antopol'skii A.B., Kalenov N.E., Serebryakov V.A., Sotnikov A.N. About the common digital space of scientific knowledge. *Bull. of the RAS*, 2019, vol. 89, no. 7, pp. 728-735 (on Russ.).

Интерактивная система терминологических словарей как один из элементов онтологии научных знаний

Н.Е. Каленов [1], *профессор, д.т.н, гл. науч. сотр. nekalenov@yandex.ru*
А.М. Сенько [2], *науч. сотрудник, alexander.senko@gmail.com*

[1] *МСЦ РАН, филиал ФГУ ФНЦ НИИСИ РАН, Москва, 119334, Россия*
[2] *БЕН РАН, Москва, 119991, Россия*

Рассматривается интерактивная информационная система, представленная в Интернете, содержащая более 12 000 терминов, относящихся к 69 тематическим разделам (словарям) науки и техники, соответствующим верхнему уровню Государственного рубрикатора научно-технической информации. Каждый термин представлен своими определениями (возможно, несколькими) и ссылками на источники определений. В систему введены рубрикаторы универсальной десятичной классификации и библиотечно-библиографической классификации (частично), а также определено соответствие индексов этих классификационных систем терминам, включенным в словари. Кроме того, представлены связи между терминами как внутри одного словаря, так и с терминами других словарей.

Система ориентирована на централизованную поддержку и распределенное редактирование и пополнение данных. В системе предусмотрены три уровня прав доступа: 1) пользовательский с правами поиска данных, их просмотра и развитой навигации; 2) редакторский, в котором к пользовательским правам добавлены права ввода и редактирования данных; 3) администраторский, где к правам редактора добавлена возможность ввода новых пользователей и установление их прав, модификация структуры базы данных. Система позволяет просматривать истории изменения информации (кто, что и когда изменял внутри контента), а также обеспечивает развитую навигацию между собственными элементами.

Дальнейшее развитие системы направлено на ее интеграцию в разрабатываемую при поддержке РФФИ модель единого цифрового пространства научных знаний в качестве основы для формирования онтологии научных направлений, входящей в ядро этого пространства.

Ключевые слова: *терминологические словари, классификационные системы, онтология научных знаний, рубрикаторы, научное пространство, научные базы данных, интерактивная система, Интернет.*

Литература

1. Белоозеров В.Н. Технология разработки терминологических словарей по лексике классификационных систем // Информационное обеспечение науки: новые технологии: сб. науч. тр. 2015. С. 126-136.

2. Антошкова О.А., Белоозеров В.Н., Дмитриева Е.Ю., Шапкин А.В. Разработка онтологии НТИ на основе библиографических классификаций Информационное обеспечение науки: новые технологии: сб. науч. тр. 2017. С. 292-300.

3. Антопольский А.Б., Белоозеров В.Н., Маркарова Т.С. О разработке онтологии на основе классификаторов научной информации и терминологических словарей // Информационные ресурсы России. 2017. № 5. С. 2-7.

4. Antopol'skii A.B., Beloozerov V.N., Kalenov N., Shaburova N.N., Yakshin M.M. The Development of a Semantic Network of Keywords Based on Definitive Relationships // Scientific and Technical Information Processing, 2017, vol. 44, no. 4, pp. 261-265. DOI: 10.3103/S0147688217040062.

5. Антопольский А.Б., Белоозеров В.Н., Каленов Н.Е., Маркарова Т.С. О развитии терминологической базы данных в виде комплекса отраслевых информационно-поисковых тезаурусов // Информационные ресурсы России. 2018. № 5. С. 22-30.

6. Государственный рубрикатор научно-технической информации. URL: http://www.extech.ru/info/catalogs/grnti/ (дата обращения: 11.06.2019).

7. Якшин М.М. Развитие платформы SciRus // Информационное обеспечение науки: новые технологии: сб. науч. тр. 2015. С. 203-207.

8. Якшин М.М. Платформа SciRus - основа технологического комплекса электронной библиотеки "Научное наследие России" // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: сб. тр. XVI Всерос. науч. конф. (RCDL-2014). 2014. С. 362-368.

9. Ruby on Rails 4.1. Release Notes. URL: https://edgeguides.rubyonrails.org/4_1_release_notes.html (дата обращения: 18.06.2019).

10. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. О едином цифровом пространстве научных знаний // Вестн. РАН. 2019. Т. 89. № 7. С. 728-735.